DOCKET NO: 282711US8X

IN THE UNITED STATES PATENT & TRADEMARK OFFICE

IN RE APPLICATION OF	:
THOMAS KEMP, ET AL.	: EXAMINER: GODBOLD, DOUGLAS
SERIAL NO: 10/731,929	:
FILED: DECEMBER 10, 2003	: GROUP ART UNIT: 2626
FOR: METHOD FOR PROCESSING SPEECH USING ABSOLUTE LOUDNESS	:

APPEAL BRIEF

COMMISSIONER FOR PATENTS
ALEXANDRIA, VIRGINIA 22313

SIR:

This Appeal Brief is filed in response to the Notification of Non-Compliant Appeal Brief under 37 C.F.R. § 41.37 of May 13, 2009, and replaces the Appeal Brief filed on April 15, 2009. This is an appeal from the fourth rejection of the claims contained in the final Office Action mailed on October 22, 2008. A Notice of Appeal was timely filed on January 22, 2009.

I. REAL PARTY IN INTEREST

The real party in interest for this appeal in the present application is Sony Deutschland GmbH, having the principal place of business in 50829 Köln, Germany, by way of Assignment recorded in the U.S. Patent and Trademark Office at Reel 017746, Frame 0583. Sony Deutschland GmbH is owned by Sony Corporation, having its principal place of business in Minato, Tokyo, Japan.

II. RELATED APPEALS AND INTERFERENCES

To the best of Appellants' knowledge there are no other appeals or interferences which will directly affect or be directly affected by, or have a bearing on, the Board's decision in this appeal.

III. STATUS OF CLAIMS

Claims 1-2, 4-9, and 12-15 are pending in this application. Claims 3 and 10-11 were cancelled by amendments during prosecution. Claims 1-2, 4-9, and 12-15 were rejected by the final Office Action of October 22, 2008. The present Appeal Brief appeals the final rejections of Claims 1-2, 4-9, and 12-15.

IV. STATUS OF AMENDMENTS

In response to a non-final Office Action of March 17, 2008, a personal interview was held between Examiners Godbold and Smits, and Appellants' representative Nikolaus P. Schibli, Ph.D., Reg. No. 56,994, on June 24, 2008. Subsequently, an Amendment was filed under 37 C.F.R. § 1.111 with amendments to independent Claims 1, 9 and 12-15. In response, the USPTO issued a final Office Action, finally rejecting Claims 1-2, 4-9, and 12-15. No amendments were filed after final. On January 22, 2009, a Notice of Appeal was timely filed.

V. SUMMARY OF THE CLAIMED SUBJECT MATTER

The claimed invention relates to a method for processing speech (Claim 1), a speech processing system (Claim 9), a computer readable medium encoded with a computer program configured to cause a processor-based device to execute a method (Claim 12), a method for

processing speech (Claim 13), a system for emotion recognition and/or speaker identification (Claim 14), and a method for processing speech (Claim 15).

By way of example as shown in Appellants' Figs. 1 and 2, a method for processing speech is provided, as recited in independent Claim 1. The method includes a step of receiving a speech input of a speaker, (see specification, p. 4, ll. 21-22, p. 5, ll. 20-21, Fig. 1, speech input SI, microphone array MA, Fig. 2, speaker "S"), generating speech parameters from said speech input, (see specification, p. 4, ll. 21-25, Fig. 1, speech parameters "SP"), determining parameters describing an absolute loudness of said speech input, the absolute loudness being a loudness of the speech at a location of a source of the speech, (see specification, p. 4, ll. 27-35, p. 5, ll. 1-14, p. 5, ll. 20-24, Fig. 1, compute distance CD, distance D, compute Loudness CL, loudness L), and evaluating at least one of said speech input and said speech parameters using said parameters describing the absolute loudness to identify the speaker. (See specification, p. 5, ll. 15-19, and from p. 5, l. 33, to. p. 6, l. 2, Fig. 1, speaker identification and/or emotion recognition EV.)

In addition, by way of example as shown in Appellants' Figs. 1 and 2, a speech processing system is provided, as recited in independent Claim 9. The speech processing system is configured to receive a speech input of a speaker, (see specification, p. 4, ll. 21-22, p. 5, ll. 20-21, Fig. 1, speech input SI, microphone array MA, Fig. 2, speaker "S"), generate speech parameters from said speech input, (see specification, p. 4, ll. 21-25, Fig. 1, speech parameters "SP"), determine parameters describing an absolute loudness of said speech input, the absolute loudness being a loudness of the speech at a location of a source of the speech, (see specification, p. 4, ll. 27-35, p. 5, ll. 1-14, p. 5, ll. 20-24, Fig. 1, compute distance CD, distance D, compute loudness CL, loudness L), and evaluate at least one of said speech input and said speech parameters using said parameters describing the absolute loudness to identify

the speaker. (See specification, p. 5, ll. 15-19, and from p. 5, l. 33, to. p. 6, l. 2, Fig. 1, speaker identification and/or emotion recognition EV.)

Moreover, by way of example as discussed in the specification at page 4, lines 5-13, a computer readable medium encoded with a computer program configured to cause a processor-based device to execute a method is provided, as recited in independent Claim 12. The executed method of Claim 12 includes the same steps of receiving, generating, determining, and evaluating, as recited in Appellants' method Claim 1, and the corresponding support in Appellants' disclosure is discussed above with reference to independent Claim 1.

Furthermore, by way of example as shown in Appellants' Figs. 1 and 2, a method for processing speech is provided, as recited in independent Claim 13. The method includes a step of receiving a speech signal of a speaker, (see specification, p. 4, ll. 21-22, p. 5, ll. 20-21, Fig. 1, speech input SI, microphone array MA, Fig. 2, speaker "S"), generating speech parameters from said speech signal, (see specification, p. 4, ll. 22-25, p. 5, ll. 15-19, Fig. 1, speech parameters "SP"), determining a distance of the speaker based on a time delay of a respective arrival of said speech signal at two or more microphones, (see specification, p. 4, ll. 27-30, Fig. 1, compute distance CD, distance D, Fig. 2), normalizing a measured loudness or energy by said distance, (see specification, p. 3, ll. 30-33, p. 4, ll. 33-35, p. 5, ll. 1-11), calculating an absolute loudness being a loudness of a speech that generated the speech signal at a location of a source of the speech, (see specification, p. 4, ll. 32-35, p. 5, ll. 1-14, p. 5, ll. 20-24, Fig. 1, compute loudness CL, loudness L), and evaluating at least one of said speech signal and said speech parameters using the normalized loudness or energy to identify the speaker. (See specification, p. 5, ll. 15-19, and from p. 5, l. 33, to. p. 6, l. 2, Fig. 1, speaker identification and/or emotion recognition EV.)

In addition, a system for emotion recognition and/or speaker identification is provided, as recited in independent Claim 14. The system includes at least two microphones

configured to receive a speech signal, (see specification, p. 4, ll. 21-22, p. 5, ll. 20-21, Fig. 1, speech input SI, microphone array MA, Fig. 2, speaker “S”), a data processor configured to generate speech parameters from said speech signal, (see specification, p. 4, ll. 22-25, Fig. 1, speech parameters SP), to determine a distance of the speaker based on a time delay of a respective arrival of said speech signal at said microphone, (see specification, p. 3, ll. 19-20, and ll. 23-25, p. 4, ll. 27-30, Fig. 1, compute distance CD, distance D, Fig. 2), to normalize a measured loudness or energy by said distance, (see specification, p. 3, ll. 30-33, p. 4, ll. 33-35, p. 5, ll. 1-11), to calculate an absolute loudness being a loudness of a speech that generated the speech signal at a location of a source of the speech, (see specification, p. 4, ll. 32-35, p. 5, ll. 1-14, p. 5, ll. 20-24, Fig. 1, compute loudness CL, loudness L), and configured to evaluate at least one of said speech signal and said speech parameters using the normalized loudness or energy to identify the speaker. (See specification, p. 5, ll. 15-19, and from p. 5, l. 33, to. p. 6, l. 2, Fig. 1, speaker identification and/or emotion recognition EV.)

Moreover, a method for processing speech is provided, as recited in independent Claim 15. The method includes the steps of receiving a speech signal of a speaker, (see specification, p. 4, ll. 21-22, p. 5, ll. 20-21, Fig. 1, speech input SI, microphone array MA, Fig. 2, speaker “S”), calculating an absolute loudness being a loudness of a speech that is generated by the speaker at a location of a source of the speech, (see specification, p. 4, ll. 27-35, p. 5, ll. 1-14, p. 5, ll. 20-24, Fig. 1, compute distance CD, distance D, compute loudness CL, loudness L), determining features from the speech signal, wherein the features are at least partly based on the absolute loudness, (see specification, p. 5, ll. 15-19, ll. 33-35), and determining an identity of the speaker based on the features. (See specification, p. 5, ll. 15-19, and from p. 5, l. 33, to. p. 6, l. 2, Fig. 1, speaker identification and/or emotion recognition EV).

In accordance with one of the features of the invention, Appellants have recognized substantial advantages to identify speakers and to identify the speaker’s state of emotion,

when information of the absolute loudness at the source of the speech can be detected. (Se specification, p. 3, ll. 7-12, and ll. 29-33.)

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

1) The first ground of rejection to be reviewed on appeal is of Claims 1, 4-9, and 12 14 under 35 U.S.C. § 103(a) over Gable et al. (U.S. Patent Application Publication No. 2005/0060153, hereinafter "Gable") in view of Brandstein et al. (Publication of the Journal of the Acoustical Society of America (JASA), "microphone-array localization error estimation with application to sensor placement," 1996, Vol. 99, No. 6, pp. 3807-3816, hereinafter "Brandstein").

2) The second ground of rejection to be reviewed on appeal is of Claim 2 under 35 U.S.C. § 103(a) over Gable in view of Brandstein, and further in view of Lee et al. (IEEE Publication from the Automatic Speech Recognition and Understanding (ASRU), "Recognition of negative emotions from the speech signal," 2001, pp. 240-243, hereinafter "Lee"). Claim 2 depends from independent Claim 1.

VII. ARGUMENT

i) THE APPLIED PRIOR ART REFERENCES, TAKEN IN ANY PROPER COMBINATION, DO NOT TEACH ALL THE FEATURES OF APPELLANTS' INDEPENDENT CLAIM 13

Briefly recapitulating, Appellants' Claim 13 is directed to a method for processing speech. The method includes the steps of receiving a speech signal of a speaker, generating speech parameters from said speech signal, determining a distance of the speaker based on a time delay of a respective arrival of said speech signal at two or more microphones, normalizing a measured loudness or energy by said distance, ***calculating an absolute***

loudness being a loudness of a speech that generated the speech signal at a location of a source of the speech, and evaluating at least one of said speech signal and said speech parameters using the normalized loudness or energy to identify the speaker.

The applied references, Gable and Brandstein, taken in any combination, fail to teach a step of calculating an absolute loudness being a loudness of a speech that generated the speech signal at a location of a source of the speech, as required by Appellants' Claim 13.

The reference Gable is directed to a system for speech characterization, where a speaker can be verified by collecting voice data and extracting parameters from his voice data and especially by collecting non-acoustic data, such as a non-acoustic glottal wave data by the use of a glottal electromagnetic micro-power sensor. (Gable, Abstract, p. 2, ¶ [0026], ll. 1-7.) Gable explains that parameters are extracted from acoustic data and non-acoustic EM data form a set of feature vectors used to calculate a performance. (Gable, p. 2, ¶¶ [0026]-[0028].) Gable explains that verification parameters for the speaker's identity may contain information on amplitude of the speech. (Gable, p. 2, ¶¶ [0027], ll. 5-8.) However, Gable fails to teach a step of calculating an absolute loudness being a loudness of a speech that generated the speech signal at a location of a source of the speech, as required by Appellants' independent Claim 13. Gable does not specify which "amplitude" of the speech he is referring to. This is also partially confirmed by the final Office Action. (October 22, 2008 Office Action, p. 11, ll. 10-15.)

However, the pending Office Action rejects the features of Appellants' independent Claim 13 by pointing out to the reference Brandstein at Sections 2 and 3 where a speaker source location problem and the estimation thereof are discussed, and in Section 5.1 where a source model is given, and also assumes that the combination of Gable and Brandstein is proper. (October 22, 2008 Office Action, p. 11, l. 16, to p. 12, l. 1-19.) Appellants respectfully disagree with these assertions.

The reference Brandstein is directed to a method capable of predicting an error region associated with a speech-source location that is obtained by a set of microphones. (Brandstein, Abstract.) Brandstein explains that his teachings can locate a source of speech by using a time-difference-of arrival analysis (TDOA) on several microphone channels. (Brandstein, p. 3, starting at l. 19.) His main goal is to detect and track a moving audio source inside a reception area, for example clearly locate a speaker in a room, to attenuate other speakers in the same room. (Brandstein, p. 1, ll. 11-12, ll. 19-21, see also p. 20, Fig. 5, showing multiple participants for a videoconference in a room.) But Brandstein is silent on a step of ***calculating an absolute loudness*** being a loudness of a speech that generated the speech signal at a location of a source of the speech, as required by Appellants' independent Claim 13.

With respect to the teachings in Brandstein at pages 3-10 and 21, the pending Office Action asserts that Brandstein "provides a relationship of a source amplitude as a function of distance and angle form [sic] the source." (Office Action, from p. 11, l. 21, to p. 12, l. 1.) In addition, the Office Action explains that Brandstein "determines a distance of the speaker based on a time delay of a respective arrival of said speech signal at two or more microphones." (Office Action, p. 11, ll. 16-20.) Finally, the Office Action concludes that the step of calculating of an absolute loudness, as required by Appellants' Claim 13, is taught by Brandstein, by asserting "[w]hen this information is combined with the source locating algorithms of section 2, ***one can obviously estimate*** the amplitude at the source itself given the amplitude at the microphone array." (Office Action, p. 12, ll. 10-13, emphasis added.)

Appellants disagree with these assertions of the final Office Action on pages 11-12. Although the mathematical model for the source of Brandstein explains that a source amplitude can depend on radiation angle and distance, (Brandstein, p. 21, ll. 10-14) Brandstein never actually calculates a loudness of a speech that generated by the speech

signal at a location of a source of the speech, as required in Appellants' Claim 13. Brandstein explains on pages 3-4 that unbiased estimates of time-difference-of-arrival (TDOA) of acoustic signals are calculated using propagation speeds and a maximum-likelihood estimation algorithm (Brandstein, from p. 4, l. 4, to p. 5, l. 3.) Similarly, in Section 3.2, related to the source estimate, Brandstein uses maximum-likelihood algorithms to calculate a geometric location of a source, using the estimated TDOAs. (Brandstein, pp. 8-10. Equation 16.) But Brandstein never calculates a source amplitude, nor does he explain how the amplitude can be calculated by giving a simple model thereof.

Regarding the Office Action's assertion that "one can obviously estimate the amplitude at the source itself given the amplitude at the microphone array," Appellants believe that this assertion of inherency of the features of Appellants' Claim 13 is insufficient to reject this claim. A mere position that a reference ***could*** perform a claimed feature is insufficient to form a rejection based on inherency. As discussed above, Brandstein does not calculate the loudness of a speech at a location of a source of the speech, but merely shows a mathematical model for source, where the amplitude is a function of radiation angle and distance. (Brandstein, p. 21, ll. 10-14.) The USPTO has the burden to show that the alleged inherent characteristic ***necessarily*** flows from the teachings of the applied references, and the USPTO has not met that burden. *See* M.P.E.P. § 2112. *See also* same section stating that "[t]he fact that a certain result or characteristic ***may*** occur or be present in the prior art is not sufficient to establish the inherency of that result or characteristic," (emphasis in original). *See also In re Robertson*, 49 USPQ2d 1949, 1951 (Fed. Cir. 1999).

Accordingly, the outstanding Office Action has not met that burden and fails to provide documentary support where such feature is taught or where such feature can be inferred from. Appellants does not challenge that Brandstein explains how the geometrical coordinates of a speech source location are estimated, but Brandstein does not teach anything

related to a step of determining parameters describing an absolute loudness, the absolute loudness being a loudness of the speech at a location of a source of the speech, as required by Appellants' Claim 13.

Therefore, even if the combination of Gable and Brandstein is assumed to be proper, the cited passages of the combination fails to teach explicitly or inherently every element of Appellants' Claim 13.

ii) THE COMBINATION OF THE APPLIED PRIOR ART REFERENCES IS NOT OBVIOUS

Appellants also respectfully submit that the combination of the references Gable and Brandstein is not obvious. The final Office Action asserted that the combination is obvious, because it would "provide a method of normalizing the loudness for speaker verification to provide a means for provide [sic] a high quality signal of the desired speaker that is not adversely effected [sic] by the distance from a speaker to the microphone array." (October 22, 2008 Office Action, p. 5, ll. 4-9).

First, from the above reasoning, it is still not clear how Brandstein's multi-speaker system with four microphones installed in a room, to identify a location of a targeted speech source, (Brandstein, Fig. 6, "microphone placements") could be incorporated into Gable's speaker verification system, having a single microphone 202 and a GEMS sensor 208 that are connected to a PC 210. (Gable, Fig. 2, p. 2, ¶ [0030].) Under such a modification, Gable's system could not be used as suggested in the pending Office Action, because in Gable's system, only one microphone 202 is used, and the speaker must approach this single microphone to make "an identity claim" during the speaker identification process 100. (Gable, p. 2, ¶ [0029].) In addition, Gable's verification system is designed to heavily rely on information from a non-acoustic sensor 208. (Gable, p. 2, ¶ [0026], Fig. 2).

Therefore, the introduction of a multi-microphone system in a room with many speakers would clearly require a substantial reconstruction or redesign of the elements of Gable where only one microphone 202 and a GEMS sensor 208 is used that is approached by a person to make a identification statement, and such redesign would change the basic principle of operation of Gable. There is no evidence that a person of ordinary skill in the art would be motivated to perform such changes and redesign. *In re Ratti*, 270 F.2d 810, 813, 123 U.S.P.Q. 349, 352 (reversing an obviousness rejection where the "suggested combination of references would require a substantial reconstruction and redesign of the elements shown in [the primary reference] as well as a change in the basic principle under which the [primary reference] construction was designed to operate.") Please note that the *In re Ratti* decision was not overruled by the Supreme Court decision of *KSR v. Teleflex*, 550 U.S. 398, 82 U.S.P.Q.2d 1385 (2007).

Second, in order to form the combination of a reference with the teachings of Gable and Brandstein, one of ordinary skill in the art would look for a reference directed to voice identification, and the estimation or calculation of absolute loudness, as required i.e. by Appellants' Claim 13.

But Gable makes it clear his system heavily relies on the information delivered by the non-acoustic GEMS sensor 208, and Brandstein is only proposing a solution to geographically locate one speaker in a room with many users of a video conferencing system based on acoustic data. (Gable, p. 2, ¶ [0026], ll. 14-15, "[t]he EM data also provides that were previously unobtainable with the all-acoustic verification systems"). Therefore, the reference Brandstein is directed to a different field of application and a different problem, but also does not teach anything related to the calculation of an absolute loudness, as discussed above with respect to subparagraph i). One of ordinary skill in the art would, therefore, not look out to the reference Brandstein to combine it with the reference Gable.

In light of the above discussion, Appellants believe that the combination of Brandstein and Gable is improper.

iii) APPELLANTS' INDEPENDENT CLAIMS 1, 9, 12, AND 14-15 ARE ALSO BELIEVED TO BE PATENTABLY DISTINCT OVER THE APPLIED PRIOR ART REFERENCES

In light of the above discussion, Appellants also respectfully submit that independent Claims 1, 9, 12, and 14-15 are also believed to be patentably distinct over the applied reference Brandstein and Gable. For example, independent Claim 1 is directed to a method for processing speech, and recites "determining parameters describing an absolute loudness of said speech input, the absolute loudness being a loudness of the speech at a location of a source of the speech." Independent Claim 9 is directed to a speech processing system, and recites that the system is configured to "determine parameters describing an absolute loudness of said speech input, the absolute loudness being a loudness of the speech at a location of a source of the speech."

Moreover, independent Claim 12 is directed to a computer readable medium encoded with a computer program configured to cause a processor-based device to execute a method, and recites a step of "determining parameters describing an absolute loudness of said speech input, the absolute loudness being a loudness of a speech at a location of a source of the speech." Independent Claim 14 is directed to a system for emotion recognition and/or speaker identification, and recites that a data processor is configured to "calculate an absolute loudness being a loudness of a speech that generated the speech signal at a location of a source of the speech." Moreover, independent Claim 15 is directed to a method for processing speech, and recites "calculating an absolute loudness being a loudness of a speech that is generated by the speaker at a location of a source of the speech."

Because independent Claims 1, 9, 12, and 14-15 include features that are analogous to the features argued above in subparagraph i) with respect to the "absolute loudness," and Claims 1, 9, 12, and 14-15 have been rejected based on an analogous obviousness rejection over the references Brandstein and Gable, Appellants respectfully submit that the arguments presented above in subparagraphs i) and ii) towards patentability of independent Claim 13 are also applicable to the patentability of independent Claims 1, 9, 12, and 14-15.

iv) APPELLANTS' DEPENDENT CLAIM 2 IS BELIEVED TO BE PATENTABLY DISTINCT OVER THE APPLIED PRIOR ART REFERENCES

Appellants request review of the second ground of rejection, and respectfully submit that the combination of Gable, Brandstein and Lee fails to teach all the features of Appellants' independent Claims 1, 9, and 12-15, even if we assume that the references can be combined, and therefore by virtue of the claim dependency, dependent Claim 2 is also believed to be allowable over these references. In addition, the feature of dependent Claim 2 are also not obvious in light of these references.

The reference Lee is directed to a method to automatically classify spoken utterances based on the emotional state of the speaker. (Lee, Abstract.) Acoustic features are calculated of the spoken utterances, such as the pitch and energy from the speech signal. (Lee, p. 241, col. 1, ll. 44-45.) The speech signals in Lee originate from a spoken dialogue over a telephone with a software machine agent implemented at a call center. (Lee, p. 241, col. 1, ll. 4-8.) The calculated acoustic features include many parameters defining the pitch and the energy level, such as mean value, median value, minimum, maximum, and range. (Lee, p. 241, col. 1, ll. 53-58.) Lee's explains that all of his samples are also normalized, that means that the origin was shifted and scaled to 1. (Lee, p. 241, col. 2, ll. 1-5.)

From the above discussion it is evident that the cited passages of Lee fail to teach a step of calculating an absolute loudness being a loudness of a speech that generated the speech signal at a location of a source of the speech, as required by Appellants' Claim 13. First, Lee uses a single microphone of a telephone for the recording, and second, Lee applies a normalization filter to all the samples. In addition, Lee clearly cites that the energy level of the speech signal ***as received at the microphone*** is calculated. Therefore, it is not possible that Lee is calculating an absolute loudness being a loudness of a speech that generated the speech signal at a location of a source of the speech, as required by Appellants' Claim 13.

Therefore, Lee fails to remedy the above argued deficiencies in subparagraphs i), ii) and iii) of Gable and/or Brandstein, even if we assume that the combination is proper. Therefore, dependent Claim 2 is also believed to be allowable by virtue of their claim dependency from independent Claim 1.

In view of the above remarks, Appellants respectfully request that the rejections of the final Office Action of October 22, 2008 be REVERSED.

Respectfully submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.

Bradley D. Lytle
Attorney of Record
Registration No. 40,073

Nikolaus P. Schibli, Ph.D.
Registered Patent Agent
Registration No. 56,994

Customer Number
22850

Tel: (703) 413-3000
Fax: (703) 413 -2220
(OSMMN 08/07)

VIII. CLAIMS APPENDIX

Claim 1: A method for processing speech, comprising:

receiving a speech input of a speaker;

generating speech parameters from said speech input;

determining parameters describing an absolute loudness of said speech input, the absolute loudness being a loudness of the speech at a location of a source of the speech; and

evaluating at least one of said speech input and said speech parameters using said parameters describing the absolute loudness to identify the speaker.

Claim 2: The method according to claim 1, wherein the step of evaluation comprises a step of emotion recognition.

Claim 3 (Cancelled).

Claim 4: The method according to claim 1, wherein a microphone array comprising a plurality of microphones is used for determining said parameters describing the absolute loudness.

Claim 5: The method according to claim 1, wherein at least one of a location and distance of the speaker is determined.

Claim 6: The method according to claim 1, wherein the absolute loudness is determined using algorithms for at least one of auditory and binaural processing.

Claim 7: The method according to claim 5, wherein

said absolute loudness is computed by normalizing a measured loudness, or energy by said distance.

Claim 8: The method according to claim 5, wherein

said distance is determined using the time delay of the speech input between said plurality of microphones.

Claim 9: A speech processing system, which is configured to:

receive a speech input of a speaker,

generate speech parameters from said speech input,

determine parameters describing an absolute loudness of said speech input, the absolute loudness being a loudness of the speech at a location of a source of the speech, and

evaluate at least one of said speech input and said speech parameters using said parameters describing the absolute loudness to identify the speaker.

Claims 10-11 (Cancelled).

Claim 12: A computer readable medium encoded with a computer program configured to cause a processor-based device to execute a method of:

receiving a speech input of a speaker,

generating speech parameters from said speech input,

determining parameters describing an absolute loudness of said speech input, the absolute loudness being a loudness of a speech at a location of a source of the speech,

evaluating at least one of said speech input and said speech parameters using said parameters describing the absolute loudness to identify the speaker.

Claim 13: A method for processing speech, comprising:

receiving a speech signal of a speaker;

generating speech parameters from said speech signal;

determining a distance of the speaker based on a time delay of a respective arrival of said speech signal at two or more microphones;

normalizing a measured loudness or energy by said distance;

calculating an absolute loudness being a loudness of a speech that generated the speech signal at a location of a source of the speech; and

evaluating at least one of said speech signal and said speech parameters using the normalized loudness or energy to identify the speaker.

Claim 14: A system for emotion recognition and/or speaker identification, comprising:

at least two microphones configured to receive a speech signal;

a data processor configured to generate speech parameters from said speech signal, to determine a distance of the speaker based on a time delay of a respective arrival of said speech signal at said microphone, to normalize a measured loudness or energy by said distance, to calculate an absolute loudness being a loudness of a speech that generated the speech signal at a location of a source of the speech; and

further configured to evaluate at least one of said speech signal and said speech parameters using the normalized loudness or energy to identify the speaker.

Claim 15: A method for processing speech comprising the steps of:

receiving a speech signal of a speaker;

calculating an absolute loudness being a loudness of a speech that is generated by the speaker at a location of a source of the speech;

determining features from the speech signal, wherein the features are at least partly based on the absolute loudness; and

determining an an identity of the speaker based on the features.

IX. EVIDENCE APPENDIX

None.

X. RELATED PROCEEDINGS APPENDIX

None.